

Digital Transformation of Korean Articulation Assessment: Reliability and Validity of Hi-DongDong

Jaewon Lee, PT, MS¹, Sejin Kwon, PT, BHS¹, Jin Young Ko, MD², Yulhyun Park, MD³, Jinju Lee, PT, MSc¹, Ju Seok Ryu, MD, PhD¹, Hojin Ju, MD¹, Jee Hyun Suh, MD, PhD^{1,*}

¹Department of Rehabilitation Medicine, Seoul National University Bundang Hospital, Seongnam, Korea

²Department of Rehabilitation, Gyeonggi Provincial Medical Center Anseong, Anseong, Korea

³Department of Rehabilitation, Gyeonggi Provincial Medical Center Icheon, Icheon, Korea

Background and Objective

- ◆ Speech sound disorders (SSD) account for ~44% of pediatric communication disorders in South Korea; articulation disorders are most prevalent
- ◆ Early detection is critical to prevent adverse effects on literacy and academic outcomes
- ◆ Urinal Test of Articulation and Phonation (UTAP): standard tool but limited by time-intensive scoring and inter-rater variability
- ◆ Hi-DongDong: app-based articulation assessment developed via 3-round Delphi (23 experts)
- ◆ Objective
 - To evaluate reliability and validity in children aged 2–13 years

Method

- ◆ Prospective multicenter study (5 institutions; Oct–Dec 2025)
- ◆ N = 931 (target >740 based on power analysis)
- ◆ Assessments:
 - Hi-DongDong (35 items)
 - UTAP-1 or UTAP-2 (30 items)
- ◆ Audio anonymized; rated by 3 SLPs
- ◆ Analyses:
 - Inter-rater reliability: ICC
 - Concurrent validity: Pearson correlation, ICC, Wilcoxon signed-rank test, Bland–Altman

Results

- ◆ **Inter-rater reliability:** Excellent (ICC \geq 0.98, $p < 0.001$)
- ◆ **Concurrent validity:**
 - UTAP-2: good agreement for PCC (ICC \geq 0.75), lower for PMLU/PVC
 - UTAP-1: excellent (PCC), moderate–good (others)
- ◆ **Correlation:**
 - Strong positive correlations with both UTAP-1 (n=337) and UTAP-2 (n=551) ($p < 0.001$)
- ◆ **Agreement analysis:**
 - Total PCC / PCC-R \rightarrow minimal bias, strong agreement
 - PMLU, PWC, PVC \rightarrow systematic differences (instrument-specific)

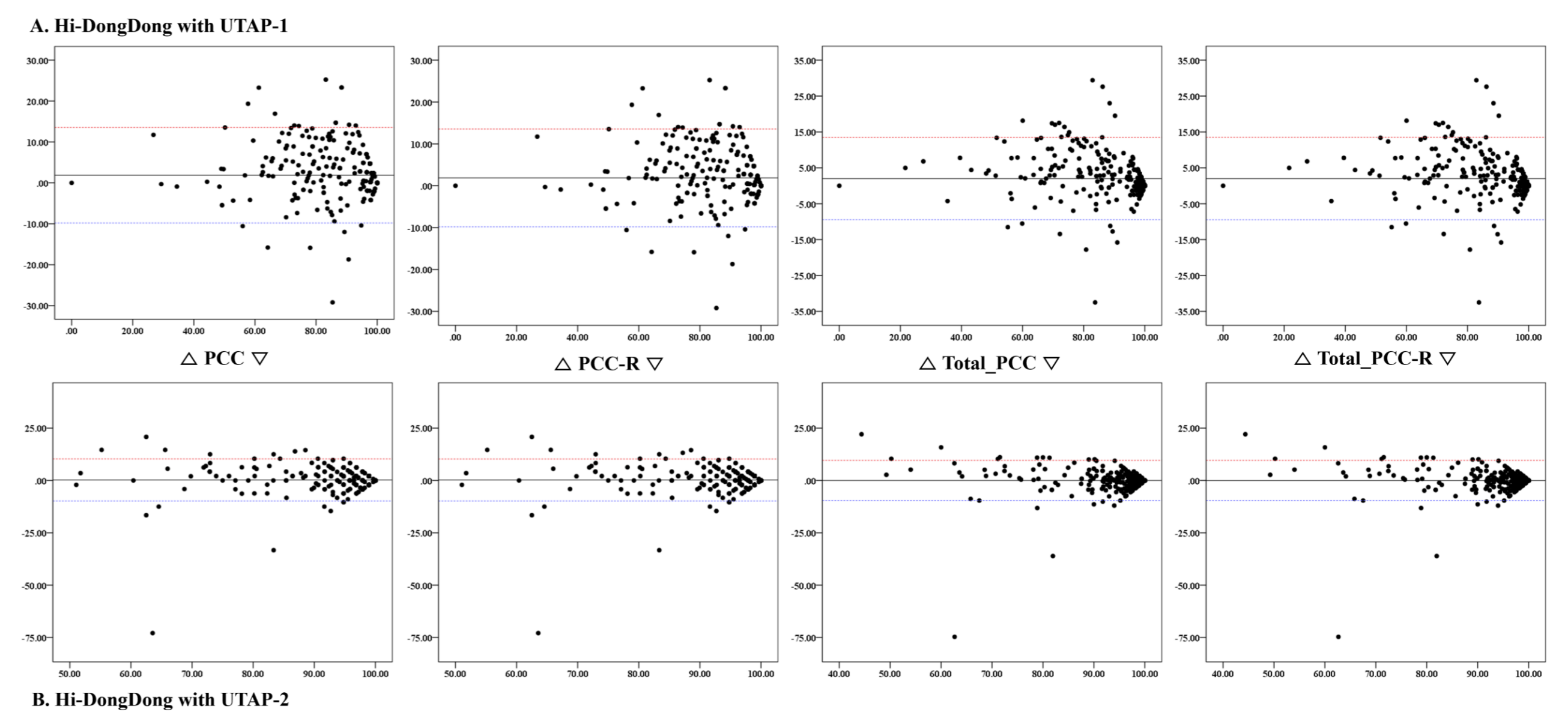


Figure 1. Bland–Altman plots comparing Hi-DongDong with UTAP-1 and UTAP-2 for PCC-based measures.

◆ Measurement stability:

- Utterance-level indices \rightarrow narrow limits of agreement
- Word-level indices \rightarrow wider limits (binary scoring effects)

Variable	Instrument	N	ICC (3, 1)	95% CI	F (df1, df2)	p
PCC	Hi-DongDong	871	0.999	[0.998, 0.999]	2060.39 (870, 1740)	0.000
	UTAP-2	551	0.998	[0.998, 0.998]	1627.41 (550, 1100)	0.000
	UTAP-1	337	0.985	[0.982, 0.988]	200.04 (336, 672)	0.000
PCC-R	Hi-DongDong	871	0.999	[0.998, 0.999]	2084.03 (870, 1740)	0.000
	UTAP-2	551	0.998	[0.998, 0.999]	1917.74 (550, 1100)	0.000
	UTAP-1	337	0.985	[0.982, 0.988]	201.99 (336, 672)	0.000
Total PCC	Hi-DongDong	871	0.999	[0.999, 0.999]	2867.72 (870, 1740)	0.000
	UTAP-2	551	0.998	[0.998, 0.999]	1899.14 (550, 1100)	0.000
	UTAP-1	337	0.983	[0.979, 0.986]	171.90 (336, 672)	0.000
Total PCC-R	Hi-DongDong	871	0.999	[0.999, 0.999]	2930.42 (870, 1740)	0.000
	UTAP-2	551	0.999	[0.999, 0.999]	2973.20 (550, 1100)	0.000
	UTAP-1	337	0.983	[0.979, 0.986]	172.59 (336, 672)	0.000
PWC	Hi-DongDong	871	0.998	[0.998, 0.998]	1729.37 (870, 1740)	0.000
	UTAP-2	551	0.998	[0.998, 0.999]	1958.90 (550, 1100)	0.000
	UTAP-1	337	0.997	[0.996, 0.997]	936.91 (336, 672)	0.000
PMLU	Hi-DongDong	871	0.999	[0.999, 0.999]	4831.67 (870, 1740)	0.000
	UTAP-2	551	0.999	[0.999, 0.999]	2387.78 (550, 1100)	0.000
	UTAP-1	337	0.987	[0.984, 0.989]	227.61 (336, 672)	0.000
PVC	Hi-DongDong	871	0.996	[0.995, 0.996]	664.87 (870, 1740)	0.000
	UTAP-2	551	0.986	[0.984, 0.988]	210.00 (550, 1100)	0.000
	UTAP-1	337	0.969	[0.962, 0.974]	93.15 (336, 672)	0.000

Table 2. Inter-rater reliability for each assessment instrument

Conclusion

- ◆ Strong psychometric performance, especially for utterance-level consonant accuracy
- ◆ Observed differences reflect analytic structure, not measurement instability
- ◆ Automated scoring improves efficiency and reduces clinician burden
- ◆ Hi-DongDong demonstrates:
 - Excellent inter-rater reliability
 - Strong concurrent validity with standardized tools
- ◆ **Utterance-level PCC indices** show highest stability
 - \rightarrow Supports clinical utility as a **reliable digital screening and monitoring tool for pediatric SSD**